CLAIM AMENDMENTS

1. - 7. (canceled)

8. (previously presented) A method of animating a synthesized model of a human face driven by an audio driving signal, comprising an analytic phase, in which

an alphabet of low level visemes is determined, and

a synthesis phase, in which

the audio driving signal is converted into a sequence of low level visemes applied to a model, wherein said analytic phase comprises the steps of

extracting both a set of information representing a shape of a speaker's face and corresponding sequences of phonetic units from a set of audio training signals;

compressing said set of information into active shape model parameter vectors representative of phonetic units;

associating to said active shape model parameter vectors representative of phonetic units an interpolation function to provide a continuous representation of movement between phonemes, wherein said interpolation function is a convex combination having combination coefficients variable as a continuous function of time whereby said association determines said alphabet of low level visemes;

associating low level parameters of facial animation, compliant with Standard ISO/IEC 14496 VER. 1, to said low level visemes;

wherein said synthesis phase comprises the steps of

extracting a sequence of phonetic units of an audio driving signal;

associating to said sequence of phonetic units extracted in said synthesis phase a corresponding sequence of low level visemes as determined in the analytic phase;

transforming said sequence of low level visemes of said synthesis phase through an interpolation function to provide a continuous representation of movement between phonemes, wherein said interpolation function of said synthesis phase is a convex combination having combination coefficients variable as a continuous function of time; and

wherein the combination coefficients carried out in the synthesis phase are the same as those used in the analytic phase.

9. (previously presented) The method according to claim 8, wherein the combination coefficients $B_n(t)$ of said convex combinations are functions of the following type:

$$\beta_n(t)=\begin{cases}\cos^2\left(\frac{\pi}{2}\frac{t-t_n}{t_{n+1}-t_n}\right); & t\in[t_n,t_{n+1}]\\ \cos^2\left(\frac{\pi}{2}\frac{t-t_n}{t_n-t_{n-1}}\right); & t\in[t_{n-1},t_n]\\ 0; & t\notin[t_{n-1},t_{n+1}]\end{cases}$$

where t_n is the instant of utterance of the nth phonetic units.

10. (previously presented) The method according to claim 9 wherein the wire-frame vertices, corresponding to model feature points, on the basis of which facial animation parameters are determined in the analytic phase, are identified and said low-level viseme interpolation operations are conducted by applying transforms on feature points for each low-level viseme, for animating a wire-frame based model.

11. (previously presented) The method according to claim 10 wherein for each position to be assumed by the model in said synthesis phase, the transforms are applied only to the vertices of the wire-frame corresponding to the feature points and the transforms are extended to remaining vertices by means of a convex combination of the transforms applied to the vertices of the wire-frame corresponding to the feature points.

12. (previously presented) The method according to claim 8 wherein said low-level visemes are converted into co-ordinates of the feature points of the face of the speaker, followed by conversion of said co-ordinates into low-level facial animation parameters compliant with Standard ISO/IEC 14496 VER.1.

13. (previously presented) The method according to claim 12 wherein said low-level facial animation parameters, representing

the co-ordinates of feature points, are obtained in the analytic phase by analyzing movements of a set of markers which identify the feature points.

14. (currently amended) The method according to claim 13 wherein data representing the co-ordinates of the feature points of the face are normalized according to the following method:

a sub-set of markers are associated to a stiff object applied to the forehead of the speaker;

the face of the speaker is set, at the beginning of the recording, to assume a position corresponding as far as possible to the position of a neutral face model, as defined in standard ISO/IEC 14496 VER. 1, and a first frame of the face in such neutral position is obtained; and

for all frames subsequent to the first frame, the sets of co-ordinates are rotated and translated so that the co-ordinates corresponding to the markers of said sub-set coincide with the co-ordinates of the markers of the same sub-set in the first frame.

15. (currently amended) A method of generating an alphabet of low level visemes for animating a synthesized model of a human face driven by an audio signal, comprising the steps of

extracting both a set of information representing the shape of a ~~speaker is~~ speaker's face and corresponding sequences of phonetic units from a set of audio training signals;

compressing said set of information into active shape model (ASM) parameter vectors; and

associating to said active shape model (ASM) parameter vectors representative of phonetic units an interpolation function to provide a continuous representation of movement between phonemes, wherein said interpolation function is a convex combination having combination coefficients variable as a continuous function of time whereby said association determines said alphabet of low level visemes.

16. (previously presented) The method according to claim 15 wherein the combination coefficients $B_n(t)$ of said convex combinations are functions of the following type:

$$\beta_n(t) = \begin{cases} \cos^2\left(\frac{\pi}{2}\frac{t-t_n}{t_{n+1}-t_n}\right); & t \in [t_n, t_{n+1}] \\ \cos^2\left(\frac{\pi}{2}\frac{t-t_n}{t_n-t_{n-1}}\right); & t \in [t_{n-1}, t_n] \\ 0; & t \notin [t_{n-1}, t_{n+1}] \end{cases}$$

where t_n is the instant of utterance of the nth phonetic units.